DOCUMENT RESUME

ED 428 080 TM 029 480

AUTHOR	Ji, Mindy F.
TITLE	A Primer on Conducting Item and Test Analyses.
PUB DATE	1999-01-23
NOTE	23p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 21-23, 1999).
PUB TYPE	Information Analyses (070) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Computation; *Computer Software; *Item Analysis; *Test Construction; Test Items

ABSTRACT

Item and test analyses can be used to revise and improve both test items and the test as a whole. Recommendations for item and test analysis practices as they are reported in commonly used measurement textbooks are summarized. A heuristic data set is used to illustrate test and item analysis practices. Techniques developed in this paper are especially useful for developing norm-referenced tests. The wide usage of computers has made conducting item and test analyses much easier than it was. Methods developed mainly for hand computation, such as the discrimination index and point-biserial correlation, should be avoided in practice, since common software can provide the necessary information. An appendix contains the Statistical Package for the Social Sciences syntax file. (Contains 3 tables and 13 references.) (SLD)

A PRIMER ON CONDUCTING ITEM AND TEST ANALYSES

Mindy F. Ji

Department of Marketing
Texas A&M University
College Station,
TX, 77843-4112
E-mail: MJI@CGSB.TAMU.EDU

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, January 23, 1999.

ABSTRACT

Item and test analyses can be employed to revise and improve both test items and the test as a whole. The purpose of this paper is to summarize the recommendations for item and test analysis practices, as these are reported in commonly-used measurement textbooks.

A general goal of test construction is to arrive at a test of minimum length that will yield scores with the necessary degree of reliability and validity for the intended uses (Crocker & Algina, 1986). In other words, the task is to develop a test composed of the best set of items (Ferketich, 1991). Item and test analyses are usually conducted to partially accomplish this goal by helping us to increase our understanding of a test, such as, why scores have specific levels of reliability and validity and how to improve these measurement characteristics (Murphy & Davidshofer, 1998; Thompson & Levitov, 1985). In addition, item and test analyses procedures allow teachers to discover items that are ambiguous, miskeyed, too easy or too difficult, and nondiscriminating (Sax, 1974). Most popular textbooks on measurement and evaluation suggest that even ordinary objective classroom tests can be improved considerably by performing item analysis (e.g., Ebel & Frisbie, 1986; Gronlund & Linn, 1990; Sax, 1974; Thorndike, Cunningham, Thorndike, & Hagen, 1991). Finally, item analysis can save the time required to develop tests that reach a given level of quality (Thompson & Levitov, 1985).

However, some best procedures in item and test analyses are too infrequently used in actual practice. The purpose of this paper is to summarize recommendations for test and item analysis practices. A concrete heuristic example is employed to illustrate how item and test analyses can be used to improve a test.

REVIEW OF TEST AND ITEM ANALYSES

Test analysis investigates the performance of all the items in a test as a set. As Thompson and Levitov (1985) noted, "In most classroom applications, test analysis focuses on the reliability of the scores generated using the test" (p. 164). Although several types of estimates of score reliability are available, Cronbach's coefficient alpha, an index of the internal consistency coefficient determined from a single administration of a test, is widely utilized. If the items are dichotomously scored, Kuder-Richardson formula #20 is equivalent to Cronbach's alpha (Reinhardt, 1996). Theoretically the reliability coefficient has a minimum value of zero and a maximum value of one, but alpha can be negative, and even less than -1 (Reinhardt, 1996). The numerator for KR_{20}, ($SD^2 - \sum pq$), actually yields the covariance terms, which involve the correlations among items. When the items are all unrelated to each other, such that the covariance is 0, KR_{20} will be 0 (Sax, 1974). Therefore a zero KR_{20} value means that each item measures something distinct from all other items; a KR_{20} value of 1, on the other hand, suggests the perfect homogeneity of items in the test.

<u>Item analysis</u> is "a term broadly used to define the computation and examination of any statistical property of examinees' responses to an individual test item" (Crocker & Algina, 1986, p. 311). The question that should be asked when examining each test item is whether an item does a good job of measuring the same thing that is measured by the rest items of the test. This question is usually answered by evaluating three factors (Murphy & Davidshofer, 1998; Thompson & Levitov, 1985). One evaluation looks at how many people answer the item correctly (item difficulty). Another aspect of item analysis is to investigate if the responses to an item are related to responses to other items on the test (item discrimination). A third aspect, which is appropriate only for certain types of items, is to examine how many people chose each response (distractor analysis). Each aspects of item analysis will be illustrated using an example later in this paper.

NORM-REFERENCED TESTS AND CRITERION-REFERENCED TESTS

The above brief discussion about item and test analyses are mainly aimed at norm-referenced tests (NRTs), which attempt to measure individual differences. Criterion-referenced tests (CRTs), in contrast, attempt to measure the attainment of some minimum level of competency (Sax, 1974). Therefore, traditional item analysis indices may not be appropriate for CRTs because most item discrimination statistics are designed to favor items on which there is substantial variation among examinees. The exact choice and interpretation of the statistics that make up an item analysis for a CRT is determined partly by the purpose of testing and partly by the person designing the analysis (Crocker & Algina, 1986). As Sax (1974) suggested, "Perhaps the best teachers can do is to construct tests that are closely tied to course objectives and to construct enough items for *each objective* to improve decision making ability" (p. 189). The following discussion will mainly focus on NRTs.

TEST AND ITEM ANALYSES USING A HEURISTIC EXAMPLE

The concepts of test and item analyses have been summarized previously. The statistical index for each concept is discussed in this section. A heuristic example will be employed to facilitate the understanding of the analyses.

Data

The first two columns in Table 1 present a hypothetical data which will be utilized in the following analyses. The example involves 30 people who have taken a 20-item multiple-choice test with five choices for each item. The number under the *Item Response* column is the frequency of examinees who marked each choice for a given item. The number of examinees who marked the right answer for each item is *bolded.* All items are dichotomously scored. Each examinee's test is scored by counting the number of answers marked correctly. The highest total score an examinee can possibly have is 20 and the lowest is 0. All the analyses were conducted using SPSS 7.5 for Windows and the syntax file presented in Appendix A.

Insert Table 1 about here

Test Statistics

Here the average total score of the 30 examinees was 9.0667 with a standard deviation of 3.4734. Table 2 is part of the output of SPSS-RELIABILITY ANALYSIS. On the bottom of the table, it shows that alpha coefficient was .7157. The relationship between alpha and each item will be discussed in details in the section of Item Statistics.

Insert Table 2 about here

Item Statistics

Item Difficulty. The most common measure of item difficulty is the proportion (or percentage) of examinees who answer the item correctly, or the *p value*. Inconsistent with the implication of its name, item difficulty actually does not provide the intrinsic characteristic of the item itself, but rather it is a behavior measure. Item difficulty is an attribute of both the item and the population taking the test (Murphy & Davidshofer, 1998). In fact, one of the limitations of classical item analysis is that item difficulty is not able to differentiate the effect of an item and the population taking the test. However, item difficulty is still a very important statistic mainly because it affects almost all total score parameters, including average item difficulty, test score mean, item variance, and total score variance (for a detailed discussion see Crocker & Algina, 1986).

Assuming a constant degree of correlation among items, items tend to improve score reliability when p_i=.50 if there is no guessing. This is because the item variance (pq) is maximized when p=.50. However, the item form of most tests (true/false, multiple-choice) allows some examinees to mark the correct response by guessing. Under the random-guessing assumption, 1/m of the portion who do not know the answer will choose the correct answer by guessing, when m is the number of choices. Therefore, items tend to improve test reliability when p_i=.50+.50/m. Our example consists of five-alternative multiple-choice items, so items tend to improve reliability when p_i=.50+.50/5=.60. The third column in Table 1 shows that items 5, 15 and 18 are very easy since everyone marked the correct answers (p=100%). Item 17 seems to be the most difficult one since only 20% of examinees answered the item correctly. Item 12 appears to be the one with the desired difficulty level (60%).

Item Discrimination. One aim of item analysis is to discover which items best measure the construct or attribute that the test is designed to measure. If the test and a single item both measure the same thing, we would expect that people who do well on the test will answer that item correctly and those who do poorly will answer that item incorrectly. In other words, a good item discriminates between those who do well on the test and those who do poorly. Three statistics can be used to measuring the discriminating power of an item: item discrimination index, the item-total correlation, and interitem correlations.

The *Discrimination Index (p)* can be only applied to dichotomously scored items. Based on the total scores of the examinees, test developers can divide examinees into two groups: people with high test scores and people with low scores. Different researchers recommend different cut scores (upper 27% and lower 27%; upper 50% and lower 50%) for this purpose. However, when sample size is reasonably large, virtually the same results can be obtained with the upper and lower 30% or 50% (Crocker & Algina, 1986). Once the upper and lower groups have been identified, the index of discrimination (D) is computed as: $D=p_u-p_l$, where p_u is the proportion in the upper group who answered the item correctly and p_l is the proportion in the lower group who answered the item correctly. Values of D range from -1.00 to 1.00. The higher the value of D, the better the item discriminates the examinees. A negative value indicates that the item inversely discriminate examinees, favoring the lower-scoring group (Crocker & Algina, 1986).

In our example, examinees were grouped into upper 50% and lower 50% since the sample size is relatively small (n=30). The Discrimination Index is listed in the fourth column of Table 1. Note that item 5, 15, and 18 have a D value of 0. The zero value implies that these items have no discrimination ability which is due to their high p values (too easy for every examinee). Ebel (1965) suggests that if $D>=.40$, the item is functioning quite satisfactorily and if $D<=.19$, the item should be eliminated or completely revised. Based on this criteria, items 2, 3, 4, 8, 10, 11, 12, 13, and 19 are good ones. Items 9, 14, and 20 should be eliminated or completely revised. The rest items may need minor or major revision.

There are several drawbacks of using D. First, D has no well known sampling distribution, therefore, it is impossible to answer questions such as how large a difference between D-values is statistically significant (Crocker & Algina, 1986). For example, it is difficult to tell if item 11

(D=46.7%) discriminates significantly better than item 12 (D=44.4%). A second shortcoming of using D is that a lot of information is lost when we convert a continuous variable, the total test score, into a dichotomous variable, upper or lower group. Finally, D does not provide any information about why an item discriminates very well or very badly. Because D can be obtained by hand computations, historically it was one of the most popular methods of reporting item discrimination effectiveness. However, it is not recommended any more because of the wide usage of computer. Today other statistics, such as item-total correlation index, should be utilized.

Item-total correlation, as its name indicates, represents a simple correlation between the score on an item and the total test score. The correlation is often referred to as a point-biserial correlation, which is a simplified computational formula for item-total correlation. Some researchers (e.g, Murphy, 1998) suggest that point-biserial correlation is just a result of lacking of computers years ago and its usage should be avoided since computers are now readily available.

The item-total correlations can be easily obtained by SPSS-RELIABILITY ANALYSIS. The column titled *Cor-r* (Corrected Item-Total Correlation) in Table 1 shows the correlation values for the items in our example. Note that items 5, 15, and 18 don't have Cor-rs since their p values are 100%. Note also that the item-total correlation is called *corrected r*, which is the correlation between an item score (i.e., 0 or 1) and the total score excluding the item. If the item is not excluded, the correlation, *uncorrected r* (Uncor-r), will appear much stronger than is warranted because the item score (e.g., 0 or 1) is one of the variables being correlated, while that score is also present within the second variable, the total score (i.e., the number of correct answers). For example, at the extreme, if the test has only one item, the uncorrected item-total correlation would necessarily always be +1. The corrected correlation coefficients should be obtained particularly when there is a small number of items in the instrument or scale (Ferketich, 1991). Table 1 lists the Cor-r and Uncor-r for each item side by side, which shows that the value of Uncor-r is larger than that of Cor-r for each item.

The item-total correlation is interpreted in much the same way as the item discrimination index, D. A positive value indicates that the item successfully discriminates between those who do relatively well on the test and those who do more poorly. An item-total correlation near zero indicates that the item does not discriminate between high and low scores. A negative value suggests that the item inversely discriminate examinees--those who do well on an item do poorly on the test.

Using item-total correlation has several advantages. First, the coefficient is just the simple correlation between an item score and the total test score. Therefore, the coefficient is easy to understand. Second, it is possible for test developers to test the statistical significance of an item-total correlation although this approach is not necessary in the practice. Third, knowing the item-total r helps us to understand the percentage of the variability (r^2) in total test scores that is accounted for by the item (Murphy & Davidshofer, 1998). Finally, item-total correlations are directly related to the reliability of test scores (Nunnally, 1982).

The last column in Table 1 listed r^2 for each item. For example, item 10 has the largest item-total correlation value (.67) and it contributes to 45% of the variability of total test scores. If this item was deleted, the alpha would decrease to .6629 from .7157 (see the last column of Table 2). This implies that item 10 is very valuable and should be retained, which is consistent with its high discrimination index value (D=83.4%). On the other hand, item 6 has the lowest Cor-r value (.05) and its contribution to the variability of total scores is near zero (Table 1). If this item was deleted, the alpha value would actually increase to .7290 (Table 2). Again this is consistent with item 6's low discrimination index value. This item probably should be removed from the test or be revised completely. The previous discussion demonstrates that item-total correlations, combined with coefficient alpha, can help test developers to detect bad items.

Although item-total correlations have many merits in item analysis, they still do not help us to understand why a particular item might show high or low levels of discriminating power. *Interritem correlations* can answer this question. Interritem correlations are the correlations among all test items. Examination of these correlations can help us to understand why some items fail to discriminate between those who do well on the test and those who do poorly. Murphy and Davidshofer (1998) suggested that if item-total correlation is low, there are two possible explanations. One possibility is that the item in question is not correlated with any of the other items on the test. In this case, this item should be rewritten or eliminated. The second possibility is that the item has positive correlations with some items, but has negative or zero correlations with other items. In this case, the test probably measures two different attributes. Table 3 is part of the correlation matrix output generated by SPSS- RELIABILITY ANALYSIS. The bolded line in the matrix assists the reader in following item 6 correlations. If we examine the correlations between item 6 (SCORE6) with other items, we can see that 10 out of 16 correlations are near zero or negative and the rest are positive. This may suggest that item 6 measures an attribute different from the one that test developer intended to measure.

Insert Table 3 about here

<u>Distractor Analysis</u>. For multiple-choice test format, another aspect of item analysis is to look at the frequency with which each incorrect response (distractor) is chosen by a group of examinees. Those who don't know the answer of an item ideally would choose randomly among all

the possible responses. In our example, 18 examinees failed to answer item 10 correctly (Table 1), we expect that each incorrect response will be chosen by four or five people (18/4=4.5).

If the number of persons choosing a distractor exceeds the random number expected (e.g., 11 examinees marked distractor B of item 10), two possibilities may occur. First, it is possible that the choice reflects partial knowledge. In this case, those examinees who scored low on the test tend to mark this distractor. A second, more troublesome possibility is that the item is a poorly constructed question. Some examinees who have more knowledge of the domain covered may be able to read into the item. For most tests, the presence of items with extremely popular decoys is likely to lower the reliability of the test scores (Murphy & Davidshofer, 1998). In other words, in this second situation, those examinees who scored high on the test will read into the test and choose a specific distractor.

In order to differentiate these two situations, we can examine the correlation between the choice of the distractor of an item and the total test score. A high negative relationship indicates the first situation, that is, the popular distractor reflects partial knowledge. A high positive relationship, however, suggests that the item is very questionable. If there is no relationship, the examinees choose the incorrect responses randomly, which is good.

As mentioned previously, item 10 seems to be a very good item (r=0.67). However, if we look at its response pattern in Table 1, we may notice that a high percent of examinees marked choice B. In order to detect if there is any troublesome relationship between marking choice B and the total score, we give those examinees who marked choice B a value of one and those who marked choice A, C, D, or E a value of zero and name this new variable WRONG10B. Then we can calculate the Pearson product-moment coefficient between WRONG10B and the total scores, which is -.339. The moderate high negative relationship indicates that those who do poorly on the test tend to mark choice B. The popularity of choice B among poor scorers probably suggests that these examinees only have partial knowledge. Therefore, we can confirm our conclusion above that item 6 performed very well. However, when the frequency of the distractor's selection is too small, this coefficient should be interpreted cautiously. In this case, the total score distribution of the examinees who marked the distractor tend to be seriously skewed, therefore, we have less confidence in calculating the correlation.

On the other hand, if examinees consistently fail to select a certain multiple choice alternative, this distractor is obviously not useful. As a result, the difficulty of the item is lowered. This decoy should be replaced or eliminated. In the case of elimination of a bad distractor, test developers should not be too concerned about the number of options used in multiple-choice format since research shows that the 3-option item is not appreciably less discriminating than the 4-option item (Crehan, Haladyna & Brewer, 1993; Trevisan, Sax & Michael, 1994).

SUMMARY

Test developers can revise and improve both test items and the test as a whole by conducting item and test analyses. The techniques discussed in this paper are especially useful for developing norm-referenced tests. Due to the wide usage of computers, conducting item and test analyses becomes much easier and more accurate than it was before. Those methods developed mainly for hand computation, such as discrimination index and point-biserial correlation, should be avoided in practice. Common software, such as SPSS, can conduct reliability analysis, which provides almost all the statistics we need in item and test analyses including the alpha coefficient, item difficulty, item-total correlation, interitem correlations, and the change in alpha coefficient if an item is deleted. When it comes to distractor analysis, test developers are encouraged to go a step further and apply the technique discussed in this paper to detect bad distractors.

REFERENCES

Crehan, K.D., Haladyna, T.M., & Brewer, B.W. (1993). Use of Inclusive Option and the Optimal Number of Options for Multiple-Choice Items. *Educational and Psychological Measurement*, *53*, 241-247.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart and Winston.

Ebel, R. L. (1965). *Measuring educational achievement*. Englewood Cliffs, NJ: Prentice-Hall.

Ebel, R. L., & Frisbie, D.A. (1986). *Essentials of educational measurement* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Ferketich, S. (1991). Focus on Psychometrics Aspects of Item Analysis. *Research in Nursing & Health, 14*, 165-168.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan.

Murphy, K. R., & Davidshofer, C.O. (1998). *Psychological testing: principles and applications.* (4th ed.). New Jersey: Prentice Hall.

Nunnally, J. C. (1982). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Reinhardt, B. (1996). Factors affecting coefficient alpha: A mini Monte Carlo study. In B. Thompson, (Ed.), *Advances in social science methodology*. (pp. 3-20). Greenwich, CT: JAI Press.

Sax, G. (1974). *Principles of educational measurement and evaluation*. Belmont, CA: Wadsworth.

Thompson, B., & Levitov, J.E. (1985). Using Microcomputers to Score and Evaluate Test Items. *Collegiate Microcomputer, 3*, 163-168.

Thorndike, R.M., Cunningham, G.K., Thorndike, R.L., & Hagen, E.P. (1991). *Measurement and evaluation in psychology and education* (5th ed.). New York: MacMillan.

Trevisan, M.S., Sax, G., & Michael, W.B. (1994). Estimating the Optimum Number of Options Per Item Using and Incremental Option Paradigm. *Educational and Psychological Measurement.*, *54 (1)*, 86-91.

Table 1

Illustrative Item Analysis Results from 30 Examinees on Items 1 to 20 of a 20-Item Test

Item	Item Response (Frequency)						Diff. p (%)	Disc Index (%)	Uncor -r	Cor-r	r^2
	A	B	C	D	E	Omit					
1	3	9	2	2	**14**	1	46.67	26.7	.35	.22	.05
2	3	**24**	2	0	1	0	80.00	40.0	.57	.48	.23
3	**20**	4	1	4	1	0	66.67	53.3	.59	.49	.24
4	5	5	2	2	**16**	0	53.33	53.3	.54	.43	.18
5	**30**	0	0	0	0	0	100.00	0.00	---	---	---
6	0	0	12	**17**	1	0	56.67	6.70	.19	.05	.00
7	4	**11**	5	4	6	0	36.67	20.0	.39	.26	.07
8	3	2	5	**15**	5	0	50.00	42.6	.35	.21	.04
9	3	10	**10**	3	4	0	33.33	13.3	.34	.21	.04
10	**12**	11	2	2	2	0	40.00	83.4	.74	.67	.45
11	9	**17**	3	0	1	0	56.67	46.7	.55	.44	.19
12	4	2	**18**	4	2	0	60.00	44.4	.61	.51	.26
13	**13**	4	2	11	0	0	43.33	46.7	.44	.31	.10
14	**14**	6	4	4	2	0	46.67	13.3	.22	.07	.00
15	0	**30**	0	0	0	0	100.00	0.00	---	---	---
16	3	3	1	**22**	1	0	73.33	26.7	.37	.25	.06
17	5	**6**	6	9	4	0	20.00	0.00	.21	.09	.01
18	0	**30**	0	0	0	0	100.00	0.00	---	---	---
19	**23**	3	1	1	2	0	76.67	46.7	.50	.39	.15
20	4	3	2	**20**	1	0	66.67	13.3	.26	.13	.02

r: Pearson Product Moment Coefficient

Table 2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
SCORE1	8.6000	11.0759	.2164	.7122
SCORE2	8.2667	10.6161	.4839	.6876
SCORE3	8.4000	10.3172	.4926	.6833
SCORE4	8.5333	10.3954	.4314	.6890
SCORE6	8.5000	11.6379	.0501	.7290
SCORE7	8.7000	10.9759	.2612	.7073
SCORE8	8.5667	11.0816	.2139	.7125
SCORE9	8.7333	11.1678	.2080	.7125
SCORE10	8.6667	9.7471	.6650	.6629
SCORE11	8.5000	10.3966	.4350	.6887
SCORE12	8.4667	10.1885	.5117	.6804
SCORE13	8.6333	10.7920	.3075	.7026
SCORE14	8.6000	11.5586	.0720	.7270
SCORE16	8.3333	11.1264	.2452	.7085
SCORE17	8.8667	11.6368	.0944	.7209
SCORE19	8.3000	10.7690	.3933	.6949
SCORE20	8.4000	11.4207	.1277	.7203

Reliability Coefficients 17 items

Alpha = .7157 **Standardized item alpha = .7170**

Table 3

Correlation Matrix

	SCORE1	SCORE2	SCORE3	SCORE4	SCORE6
SCORE1	1.0000				
SCORE2	.1336	1.0000			
SCORE3	-.0472	.5303	1.0000		
SCORE4	.2054	.3675	.1890	1.0000	
SCORE6	**.1438**	**.2354**	**.0951**	**.2607**	1.0000
SCORE7	-.1571	.2075	.3913	-.1202	**.1070**
SCORE8	.1336	.3333	.0000	.1336	**-.0673**
SCORE9	-.0945	.0000	.3500	-.1890	**-.0951**
SCORE10	.3273	.4082	.4330	.3546	**.3021**
SCORE11	.2787	.0673	.0951	.2607	**-.2217**
SCORE12	.3546	.2722	.1443	.1909	**-.1648**
SCORE13	-.1438	.1009	.0476	.4135	**.0860**
SCORE14	.0625	-.0334	.0945	.2054	**-.3955**
SCORE16	.1108	.2638	.2132	-.1108	**-.0710**
SCORE17	.0334	-.1667	.1768	.3007	**.1009**
SCORE19	.0421	.3152	.2786	.4318	**-.0053**
SCORE20	.0945	.1768	.4000	.0472	**.0951**

Appendix A

```
RECODE
 item1
 ('e'=1) ('a'=0) ('b'=0) ('c'=0) ('d'=0) (sysmis=0) INTO score1 .
EXECUTE .

RECODE
 item2
 ('b'=1) ('a'=0) ('c'=0) ('d'=0) ('e'=0) INTO score2 .
EXECUTE .

RECODE
 item3
 ('a'=1) ('b'=0) ('c'=0) ('d'=0) ('e'=0) INTO score3 .
EXECUTE .

RECODE
 item4
 ('e'=1) ('a'=0) ('b'=0) ('c'=0) ('d'=0) INTO score4 .
EXECUTE .

RECODE
 item5
 ('a'=1) ('b'=0) ('c'=0) ('d'=0) ('e'=0) INTO score5 .
EXECUTE .

RECODE
 item6
 ('d'=1) ('a'=0) ('b'=0) ('c'=0) ('e'=0) INTO score6 .
EXECUTE .

RECODE
 item7
 ('b'=1) ('a'=0) ('c'=0) ('d'=0) ('e'=0) INTO score7 .
EXECUTE .

RECODE
 item8
 ('d'=1) ('a'=0) ('b'=0) ('c'=0) ('e'=0) INTO score8 .
EXECUTE .

RECODE
 item9
 ('c'=1) ('a'=0) ('b'=0) ('d'=0) ('e'=0) INTO score9 .
EXECUTE .

RECODE
 item10
 ('a'=1) ('b'=0) ('c'=0) ('d'=0) ('e'=0) INTO score10 .
EXECUTE .

RECODE
 item11
 ('b'=1) ('a'=0) ('c'=0) ('d'=0) ('e'=0) INTO score11 .
EXECUTE .

RECODE
 item12
 ('c'=1) ('a'=0) ('b'=0) ('d'=0) ('e'=0) INTO score12 .
EXECUTE .

RECODE
 item12
 ('c'=1) ('a'=0) ('b'=0) ('d'=0) ('e'=0) INTO score12 .
EXECUTE .

RECODE
 item13
 ('a'=1) ('b'=0) ('c'=0) ('d'=0) ('e'=0) INTO score13 .
EXECUTE .

RECODE
 item14
 ('a'=1) ('b'=0) ('c'=0) ('d'=0) ('e'=0) INTO score14 .
EXECUTE .
```

```
RECODE
 item15
 ('b'=1) ('a'=0) ('c'=0) ('d'=0) ('e'=0) INTO score15 .
EXECUTE .

RECODE
 item16
 ('d'=1) ('a'=0) ('b'=0) ('c'=0) ('e'=0) INTO score16 .
EXECUTE .

RECODE
 item17
 ('b'=1) ('a'=0) ('c'=0) ('d'=0) ('e'=0) INTO score17 .
EXECUTE .

RECODE
 item18
 ('b'=1) ('a'=0) ('c'=0) ('d'=0) ('e'=0) INTO score18 .
EXECUTE .

RECODE
 item19
 ('a'=1) ('b'=0) ('c'=0) ('d'=0) ('e'=0) INTO score19 .
EXECUTE .

RECODE
 item20
 ('d'=1) ('a'=0) ('b'=0) ('c'=0) ('e'=0) INTO score20 .
EXECUTE .

COMPUTE scoretot = score1 + score2 + score3 + score4 + score5 +score6 +
 score7 + score8 + score9 + score10 + score11 + score12 + score13 + score14
 + score15 + score16 + score17 + score18 + score19 + score20 .
EXECUTE .
SORT CASES BY
 scoretot (A) .

FREQUENCIES
 VARIABLES=scoretot
 /NTILES= 4
 /PERCENTILES= 50
 /STATISTICS=STDDEV MEAN MEDIAN .
STRING group (A8) .

RECODE
 scoretot
 (5='low') (6='low') (7='low') (8='low') (9='low') (10='low')
 (11='low') (12='high') (13='high') (14='high') (15='high') (16='high')
 (17='high') (18='high') INTO group .
EXECUTE .

Report
 /FORMAT= CHWRAP(ON) BRKSPACE(-1) SUMSPACE(0) AUTOMATIC
  PREVIEW(OFF) CHALIGN(BOTTOM) CHDSPACE(1)
  UNDERSCORE(ON) ONEBREAKCOL(OFF)
 PAGE(1) MISSING'.' LENGTH(1, 59)ALIGN(LEFT) TSPACE(1) FTSPACE(1)
 MARGINS(1,82)
 /TITLE=
 RIGHT 'Page )PAGE'
 /VARIABLES
 score1 'score1' 'Sum' (RIGHT) (OFFSET(0)) (10)
 score2 'score2' 'Sum' (RIGHT) (OFFSET(0)) (10)
 score3 'score3' 'Sum' (RIGHT) (OFFSET(0)) (10)
 score4 'score4' 'Sum' (RIGHT) (OFFSET(0)) (10)
 score5 'score5' 'Sum' (RIGHT) (OFFSET(0)) (10)
 score6 'score6' 'Sum' (RIGHT) (OFFSET(0)) (10)
 score7 'score7' 'Sum' (RIGHT) (OFFSET(0)) (10)
 /BREAK (TOTAL)
 /SUMMARY SUM(score1) SUM(score2) SUM(score3) SUM(score4)
 SUM(score5) SUM(score6) SUM(score7) 'Grand Total' (1) .

CROSSTABS
 /TABLES=item1 BY group
 /FORMAT= AVALUE TABLES
 /CELLS= COUNT ROW COLUMN ....
```

```
CROSSTABS
 /TABLES=item20  BY group
 /FORMAT= AVALUE TABLES
 /CELLS= COUNT ROW COLUMN .

CROSSTABS
 /TABLES=score1  BY group
 /FORMAT= AVALUE TABLES
 /CELLS= COUNT ROW COLUMN .

CROSSTABS
 /TABLES=score20  BY group
 /FORMAT= AVALUE TABLES
 /CELLS= COUNT ROW COLUMN .

RELIABILITY variables=score1 to score20/
  scale(total)=score1 to score20/
 model=alpha/statistics=all/
 summary=total .

RECODE
 item10
 ('a'=0) ('b'=1) ('c'=0) ('d'=0) ('e'=0) INTO wrong10b
EXECUT
CORRELATIONS
 /VARIABLES=scoretot wrong 10b
 /PRINT=TWOTAL NOSIG
 /MISSING=PAIRWISE.
```

TM029480

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A PRIMER ON CONDUCTING ITEM AND TEST ANALYSIS	
Author(s): MINDY F. JI	
Corporate Source:	Publication Date: 1/99

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

☒ Sample sticker to be affixed to document | Sample sticker to be affixed to document ☐

Check here Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

MINDY F. JI

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

or here Permitting reproduction in other than paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: X Mindy Ji	Position: RES ASSOCIATE
Printed Name: MINDY F. JI	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: (409) 845-1831
	Date: 1/13/99

OVER